



**HAL**  
open science

## Characterization of a relaxase belonging to the MOB<sub>T</sub> family, a widespread family in Firmicutes mediating the transfer of ICEs

Nicolas Soler, Emilie Robert, Isaure Chauvot de Beauchêne, Philippe Monteiro, Virginie Libante, Bernard Maignet, Johan Staub, David W. Ritchie, Gérard Guédon, Sophie Payot, et al.

### ► To cite this version:

Nicolas Soler, Emilie Robert, Isaure Chauvot de Beauchêne, Philippe Monteiro, Virginie Libante, et al.. Characterization of a relaxase belonging to the MOB<sub>T</sub> family, a widespread family in Firmicutes mediating the transfer of ICEs. *Mobile DNA*, 2019, 10 (1), pp.1-16. 10.1186/s13100-019-0160-9 . hal-02138843

HAL Id: hal-02138843

<https://inria.hal.science/hal-02138843v1>

Submitted on 24 May 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH

Open Access



# Characterization of a relaxase belonging to the MOB<sub>T</sub> family, a widespread family in Firmicutes mediating the transfer of ICEs

Nicolas Soler<sup>1\*</sup> , Emilie Robert<sup>1</sup>, Isaure Chauvot de Beauchêne<sup>2</sup>, Philippe Monteiro<sup>1</sup>, Virginie Libante<sup>1</sup>, Bernard Maignet<sup>2</sup>, Johan Staub<sup>1</sup>, David W. Ritchie<sup>2</sup>, Gérard Guédon<sup>1</sup>, Sophie Payot<sup>1</sup>, Marie-Dominique Devignes<sup>2</sup> and Nathalie Leblond-Bourget<sup>1\*</sup>

## Abstract

**Background:** Conjugative spread of antibiotic resistance and virulence genes in bacteria constitutes an important threat to public health. Beyond the well-known conjugative plasmids, recent genome analyses have shown that integrative and conjugative elements (ICEs) are the most widespread conjugative elements, even if their transfer mechanism has been little studied until now. The initiator of conjugation is the relaxase, a protein catalyzing a site-specific nick on the origin of transfer (*oriT*) of the ICE. Besides canonical relaxases, recent studies revealed non-canonical ones, such as relaxases of the MOB<sub>T</sub> family that are related to rolling-circle replication proteins of the *Rep\_trans* family. MOB<sub>T</sub> relaxases are encoded by ICEs of the ICESt3/ICEBs1/Tn916 superfamily, a superfamily widespread in Firmicutes, and frequently conferring antibiotic resistance.

**Results:** Here, we present the first biochemical and structural characterization of a MOB<sub>T</sub> relaxase: the RelSt3 relaxase encoded by ICESt3 from *Streptococcus thermophilus*. We identified the *oriT* region of ICESt3 and demonstrated that RelSt3 is required for its conjugative transfer. The purified RelSt3 protein is a stable dimer that provides a Mn<sup>2+</sup>-dependent single-stranded endonuclease activity. Sequence comparisons of MOB<sub>T</sub> relaxases led to the identification of MOB<sub>T</sub> conserved motifs. These motifs, together with the construction of a 3D model of the relaxase domain of RelSt3, allowed us to determine conserved residues of the RelSt3 active site. The involvement of these residues in DNA nicking activity was demonstrated by targeted mutagenesis.

**Conclusions:** All together, this work argues in favor of MOB<sub>T</sub> being a full family of non-canonical relaxases. The biochemical and structural characterization of a MOB<sub>T</sub> member provides new insights on the molecular mechanism of conjugative transfer mediated by ICEs in Gram-positive bacteria. This could be a first step towards conceiving rational strategies to control gene transfer in these bacteria.

**Keywords:** Integrative and conjugative elements, Relaxase, MOB<sub>T</sub>, Conjugation, Tn916, Firmicutes

## Background

Horizontal gene transfer (HGT) drives the evolution of bacterial genomes and plays an important role in the adaptation of bacteria to different environments or the colonization of novel niches. Conjugative elements are major actors that trigger HGT [1, 2]. Besides conjugative plasmids, many integrative and conjugative elements (ICEs)

in the host chromosome encode their own DNA transfer from donor to recipient cells by conjugation [3–5]. Genome analyses indicate that ICEs are widespread in most phyla of bacteria, and that many of them carry adaptive genes such as antibiotic resistance or virulence genes [3, 4, 6, 7], hence contributing to the dissemination of these traits among bacterial populations. Thus, many Firmicutes have recently become resistant to antibiotics through ICE transfer, including various strains of Clostridia, Enterococci, Staphylococci and Streptococci, some of which are severe human or zoonotic pathogens [8–14]. Although Gram-positive ICEs

\* Correspondence: nicolas.soler@univ-lorraine.fr; nathalie.leblond@univ-lorraine.fr

<sup>1</sup>Université de Lorraine, Inra, UMR1128 DynAMic, F-54000 Nancy, France  
Full list of author information is available at the end of the article



are more abundant than conjugative plasmids, especially in Firmicutes genomes [7], their molecular mechanism is still poorly characterized. A recent analysis of 124 completely sequenced streptococcal genomes identified 3 major superfamilies of ICEs in streptococci, the most widespread of which being the *ICES<sub>t3</sub>/ICEBs<sub>1</sub>/Tn916* superfamily [15]. As *Tn916* and its relatives have been involved in the spread of various antimicrobial resistance genes, especially resistance to tetracyclines and macrolides [6, 14, 16], the molecular characterization of the conjugation machinery from *ICES<sub>t3</sub>/ICEBs<sub>1</sub>/Tn916* superfamily is critical.

Prior to transfer, ICEs excise from the host chromosome as a circular form. As in conjugative plasmids, the initiation of conjugative transfer is likely performed by a multi-protein complex called the relaxosome, that recognizes the origin of transfer (*oriT*) on the excised ICE DNA to be transferred [17]. As demonstrated for conjugative plasmids, the main protein of the relaxosome is the relaxase, a trans-esterase enzyme encoded by ICEs. The relaxase introduces a single-stranded nick at the *nic* site of *oriT*. Whereas the 3'-end of the nick is used to initiate rolling-circle replication of the element, the relaxase remains covalently bound to the 5'-end via a catalytic tyrosine residue [18]. A coupling protein is assumed to recruit the relaxase-DNA complex to a type-IV secretion system (T4SS) encoded by the ICE, providing energy to translocate this complex into the recipient cell [19]. The relaxase is then thought to join the transferred ssDNA at the *nic* site, leading to a re-circularization of the ssDNA in the recipient cell [20].

Relaxases, also termed Mob (for mobilization) proteins, belong to different families of proteins. The canonical relaxases harbor a histidine triad with a 'HxH' signature that is known to coordinate a metallic cofactor [21]. They include the MOB<sub>F</sub>, MOB<sub>P</sub>, MOB<sub>L</sub>, MOB<sub>Q</sub>, MOB<sub>V</sub> and MOB<sub>B</sub> families [7, 18, 22, 23]. This 'HxH' motif is also found in rolling-circle replication initiators of numerous plasmids and viruses, usually termed Rep proteins [24]. Several other families of relaxases have been identified which do not harbor the 'HxH' signature and which are unrelated to canonical relaxases [23]. The best described are the MOB<sub>H</sub> family related to HD-hydrolases [23, 25], and the MOB<sub>C</sub> family of relaxases that has no typical relaxase motif, but surprisingly has been shown to be related to restriction endonucleases [26]. Additionally, recent observations indicate that several other protein families also function as relaxases, especially in Firmicutes. This is the case of the TcpM protein from the pCW3 conjugative plasmid (*Clostridium perfringens*) that is related to tyrosine recombinases [27]. Furthermore, rolling circle replication (RCR) initiators of the *Rep\_1* family have been found to act as a relaxase, being responsible for the mobilization of plasmids by ICEs, in addition to their role in plasmid maintenance [28].

The *ICES<sub>t3</sub>/ICEBs<sub>1</sub>/Tn916* superfamily of ICEs, which is widespread in Firmicutes, also encodes non-canonical relaxases. These relaxases were unclassified by Garcillan-Barcia et al [23], but were grouped in a family called MOB<sub>T</sub> in 2011 [7]. Interestingly, putative MOB<sub>T</sub> relaxases have also been detected in a number of putative integrated and mobilizable elements (IMEs) in streptococcal genomes [29, 30]. Up to now, the activity of only two MOB<sub>T</sub> relaxases has been demonstrated, namely those encoded by *Tn916* (Orf20) and *ICEBs<sub>1</sub>* (Nick). The Orf20 protein was reported to involve the *Tn916* integrase for DNA sequence specificity [31]. However, this result needs to be confirmed because the Orf20 version used lacked the N-terminal helix-turn-helix (HTH) DNA-binding domain due to misidentification of the start codon [32]. The Nick relaxase has been shown to be essential for conjugation and the *nic* site within *ICEBs<sub>1</sub>* *oriT* was identified [33]. Up to now, no MOB<sub>T</sub> relaxase has been characterized biochemically or structurally.

MOB<sub>T</sub> relaxases are non-canonical relaxases that harbor a PF02486 domain. This domain is also found within Rolling Circle Replication (RCR) initiator proteins of the *Rep\_trans* family which are involved in the maintenance of many small Firmicutes plasmids [34]. A structural and biochemical characterization of the active site of the *Rep\_trans* RCR initiator protein encoded by the pSTK1 plasmid (*Geobacillus stearothermophilus*) has been performed recently [34].

From a structural point of view, 3D structures are available for three canonical relaxase families: MOB<sub>F</sub> [23, 35–37], MOB<sub>Q</sub> [23, 38], and MOB<sub>V</sub> [39]. To date, no 3D structure is known for the non-canonical relaxases (MOB<sub>H</sub>, MOB<sub>C</sub>, and MOB<sub>T</sub>).

In this work, we characterized the MOB<sub>T</sub> relaxase encoded by *ICES<sub>t3</sub>*, hereafter called RelSt3. RelSt3 appeared to be essential for *ICES<sub>t3</sub>* conjugation, and we determined experimentally the *oriT* sequence. The RelSt3 protein was shown to be dimeric and to exhibit the expected relaxation activity that is dependent on a metallic cation, preferentially Mn<sup>2+</sup>. We also investigated the phylogenetic relationships between MOB<sub>T</sub> and *Rep\_trans* proteins, and we propose a 3D model of RelSt3 by homology with the RepSTK1 structure. Critical residues conserved within *Rep\_trans* proteins were mutagenized in RelSt3, and we demonstrated their involvement in RelSt3 activity. Our results constitute the first biochemical characterization of a MOB<sub>T</sub> relaxase.

## Results

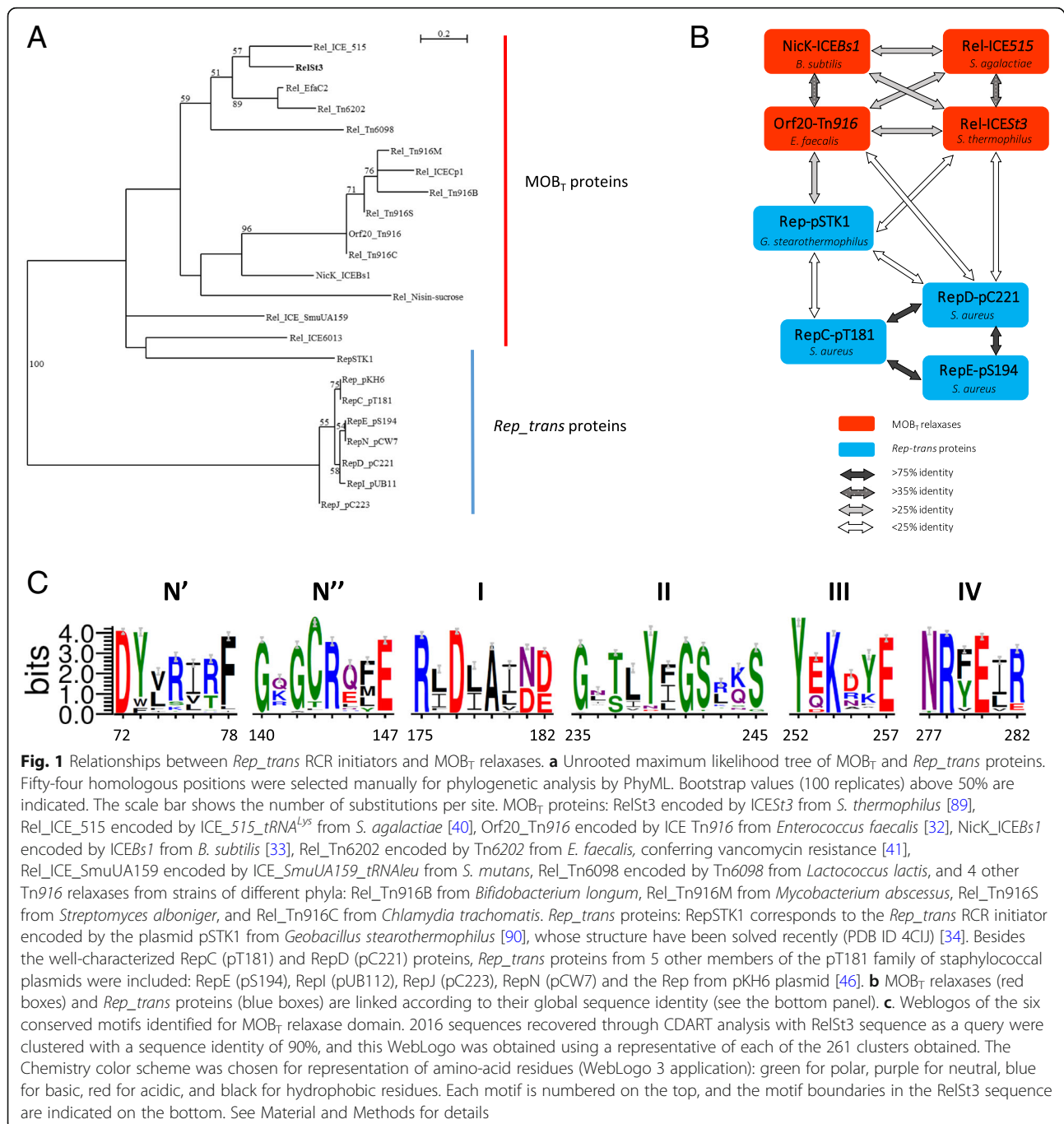
### RelSt3 is a MOB<sub>T</sub> relaxase related to *Rep\_trans* RCR initiators

The relationship between MOB<sub>T</sub> relaxases and *Rep\_trans* RCR initiators encoded by small plasmids of the pT181 family has been reported previously [33, 34].

To investigate further this relationship, a selected set of MOB<sub>T</sub> relaxase sequences including RelSt3 was phylogenetically analyzed together with *Rep\_trans* RCR initiator sequences from staphylococcal plasmids (Fig. 1a). The MOB<sub>T</sub> sequences selected for this analysis include relaxases encoded by the characterized ICEs ICEBs1 (NicK), Tn916 (Orf20), ICE<sub>515</sub>*tRNA*<sup>Lys</sup>, ICECp1, ICE6013, Tn6202, Tn6098, and nisin-sucrose transposon [31, 33, 40–45], and putative relaxases encoded by several Tn916-related elements from different bacterial

phyla, and by other distantly related ICEs belonging to the ICESt3/ICEBs1/Tn916 superfamily. *Rep\_trans* sequences include RepSTK1 whose 3-D structure was recently solved [34], and several members of the Rep proteins belonging to the pT181 family [46]. As expected, RelSt3 groups with the MOB<sub>T</sub> proteins (Fig. 1a).

The topology of the tree clearly indicated that MOB<sub>T</sub> and *Rep\_trans* from the pT181 family are distantly related (Fig. 1a). Interestingly, the RepSTK1 sequence grouped with MOB<sub>T</sub> sequences rather than with other



**Fig. 1** Relationships between *Rep\_trans* RCR initiators and MOB<sub>T</sub> relaxases. **a** Unrooted maximum likelihood tree of MOB<sub>T</sub> and *Rep\_trans* proteins. Fifty-four homologous positions were selected manually for phylogenetic analysis by PhyML. Bootstrap values (100 replicates) above 50% are indicated. The scale bar shows the number of substitutions per site. MOB<sub>T</sub> proteins: RelSt3 encoded by ICESt3 from *S. thermophilus* [89], Rel\_ICE\_515 encoded by ICE<sub>515</sub>*tRNA*<sup>Lys</sup> from *S. agalactiae* [40], Orf20\_Tn916 encoded by ICE Tn916 from *Enterococcus faecalis* [32], NicK\_ICEBs1 encoded by ICEBs1 from *B. subtilis* [33], Rel\_Tn6202 encoded by Tn6202 from *E. faecalis*, conferring vancomycin resistance [41], Rel\_ICE\_SmuUA159 encoded by ICE<sub>SmuUA159</sub>*tRNA*<sub>Leu</sub> from *S. mutans*, Rel\_Tn6098 encoded by Tn6098 from *Lactococcus lactis*, and 4 other Tn916 relaxases from strains of different phyla: Rel\_Tn916B from *Bifidobacterium longum*, Rel\_Tn916M from *Mycobacterium abscessus*, Rel\_Tn916S from *Streptomyces alboniger*, and Rel\_Tn916C from *Chlamydia trachomatis*. *Rep\_trans* proteins: RepSTK1 corresponds to the *Rep\_trans* RCR initiator encoded by the plasmid pSTK1 from *Geobacillus stearothermophilus* [90], whose structure have been solved recently (PDB ID 4CIJ) [34]. Besides the well-characterized RepC (pT181) and RepD (pC221) proteins, *Rep\_trans* proteins from 5 other members of the pT181 family of staphylococcal plasmids were included: RepE (pS194), RepI (pUB112), RepJ (pC223), RepN (pCW7) and the Rep from pKH6 plasmid [46]. **b** MOB<sub>T</sub> relaxases (red boxes) and *Rep\_trans* proteins (blue boxes) are linked according to their global sequence identity (see the bottom panel). **c** Weblogos of the six conserved motifs identified for MOB<sub>T</sub> relaxase domain. 2016 sequences recovered through CDART analysis with RelSt3 sequence as a query were clustered with a sequence identity of 90%, and this WebLogo was obtained using a representative of each of the 261 clusters obtained. The Chemistry color scheme was chosen for representation of amino-acid residues (WebLogo 3 application): green for polar, purple for neutral, blue for basic, red for acidic, and black for hydrophobic residues. Each motif is numbered on the top, and the motif boundaries in the RelSt3 sequence are indicated on the bottom. See Material and Methods for details

*Rep\_trans* proteins, even if this topology was not clearly supported by bootstrap analysis. This relationship between RepSTK1 and MOB<sub>T</sub> proteins was also supported by sequence-based searches for fold similarities using the Phyre2 [47] and iTasser [48] programs. Indeed, the use of any MOB<sub>T</sub> sequence as query gave the RepSTK1 structure as first hit, with the RepDE structure that belongs to the pT181 family being the second hit [34]. A comparison of sequence identities also illustrates that RepSTK1 is closer to MOB<sub>T</sub> relaxases than the pT181 cluster (Fig. 1b). As expected, relaxases encoded by Tn916 and Tn916-like elements clustered together among MOB<sub>T</sub> proteins with a good bootstrap support. RelSt3 is more closely related to the relaxase encoded by ICE<sub>515</sub>*tRNA<sup>Lys</sup>* from *S. agalactiae* than to Nick encoded by ICEBs1 from *B. subtilis*. These relationships were also confirmed by multiple alignments of MOB<sub>T</sub> and *Rep\_trans* protein sequences (Additional files 1 and 2: Figure S1 and S2), which allowed the identification in RelSt3 and other MOB<sub>T</sub> relaxases of the 5 conserved motifs described in *Rep\_trans* proteins by Carr et al [34]. Taken together, these data argue in favor of RelSt3 being a bona fide MOB<sub>T</sub> relaxase.

To identify conserved motifs among MOB<sub>T</sub> proteins, we recovered 2016 sequences of proteins harboring the same architecture as RelSt3, i.e. with a N-terminal HTH domain and a PF02486 domain (see Materials and Methods for details). To our knowledge, this architecture is only found in proteins encoded by ICEs (and not by plasmids), suggesting that they correspond to relaxases involved in conjugation and not to RCR *Rep\_trans* proteins. Multiple sequence alignment followed by WebLogo analysis allowed the identification of 6 conserved motifs (Fig. 1c), distinct from all other motifs found in the other MOB families [22, 23]. Among these motifs, 5 have already been identified in *Rep\_trans* proteins, whereas an additional motif, called N'', was found between motifs N' and I (Fig. 1c). The 3 acidic residues involved in cation coordination in *Rep\_trans* proteins [34] were also conserved in the same 3 motifs in MOB<sub>T</sub> relaxases: motifs N' (D72 in RelSt3 sequence), I (D177) and IV (D280). In motif III, the catalytic tyrosine identified in *Rep\_trans* proteins [49] is also conserved (Y252 in RelSt3 sequence).

#### The RelSt3 gene (*orfI*) is required for ICESt3 conjugation

If it encodes a functional relaxase, the *orfI* gene is expected to be essential for ICESt3 conjugative transfer, as demonstrated for the related gene *nick* from ICEBs1 [33]. We constructed a *S. thermophilus* strain carrying a derivative of ICESt3cat deleted for the *orfI* gene (referred as LMG18311 ICESt3Δ*orfI*cat, see Additional file 4, Supplementary Materials and methods). The ICESt3 conjugation/recombination modules are composed of 17

genes transcriptionally coupled, including *orfI* at the fifth position downstream of the P<sub>cr</sub> promoter [50]. Our strategy was to delete the *orfI* CDS while keeping the predicted *oriT* region and other ICESt3cat genes intact, without introducing any foreign sequence. The mating frequency for the LMG18311 ICESt3cat [51] as donor strain was  $1.0 \times 10^{-4} \pm 2.0 \times 10^{-5}$  transconjugants per donor cell. No transconjugant was observed when three independent clones of LMG18311 ICESt3Δ*orfI*cat were used (with a detection level  $\geq 10^{-9}$  transconjugants per donor cell), thus demonstrating that the *orfI* gene is required for ICESt3 conjugation.

#### Determination of ICESt3 *oriT*

In order to determine the location of ICESt3 *oriT*, we constructed the pOri1180-*oriT* plasmid harboring the circa 200 bp sequence of the intergenic *orfK/orfI* region from ICESt3, where the conserved *nic* site was previously predicted [52]. This plasmid carrying a spectinomycin resistance marker was tested for mobilization *in trans* by ICESt3, and was successfully transferred to recipient cells with a mobilization frequency of  $2.5 \times 10^{-3}$ . In contrast, no transconjugant was obtained with the plasmid lacking the *orfK-orfI* intergenic region (below the  $10^{-6}$  detection level) (Table 1). These results confirmed that the *oriT* sequence is located between *orfK* and *orfI* genes in ICESt3 sequence, and indicated that RelSt3 is able to mobilize a replicon *in trans*.

#### RelSt3 is dimeric in solution

RelSt3 was over-expressed and purified from *E. coli* using affinity (nickel) chromatography (Fig. 2a) as described in Additional file 4, Supplementary Materials and Methods. RelSt3 was purified to homogeneity, with a yield of about 1.5 mg per g of cell pellet. As the final gel filtration step was performed with a calibrated Superdex 200 column (Additional file 3, Figure S3), the elution volume of the single elution peak (72.4 mL) allowed us to estimate the apparent molecular weight of the protein in solution to be about 88 kDa (Fig. 2b). As the predicted size of RelSt3 is 48.5 kDa, this finding suggested that RelSt3 is dimeric.

The RelSt3 protein was also submitted to chemical cross-linking experiments. To do so, RelSt3 was incubated with increasing concentrations of glutaraldehyde (0.001 to 0.1%, Fig. 2c), and a well-defined band appeared with 0.01% glutaraldehyde around 100 kDa, a molecular mass corresponding to twice the predicted mass of the monomer. At higher concentration of cross-linker, higher molecular mass species appeared.

We next studied the protein homogeneity by velocity sedimentation in order to determine the molecular weight and radius of RelSt3 in solution (Fig. 2d). Data were collected using 1 and 5 μM of protein, which both gave



**Table 1** ICES<sub>3</sub> mobilization *in trans* of an *oriT*-harboring plasmid

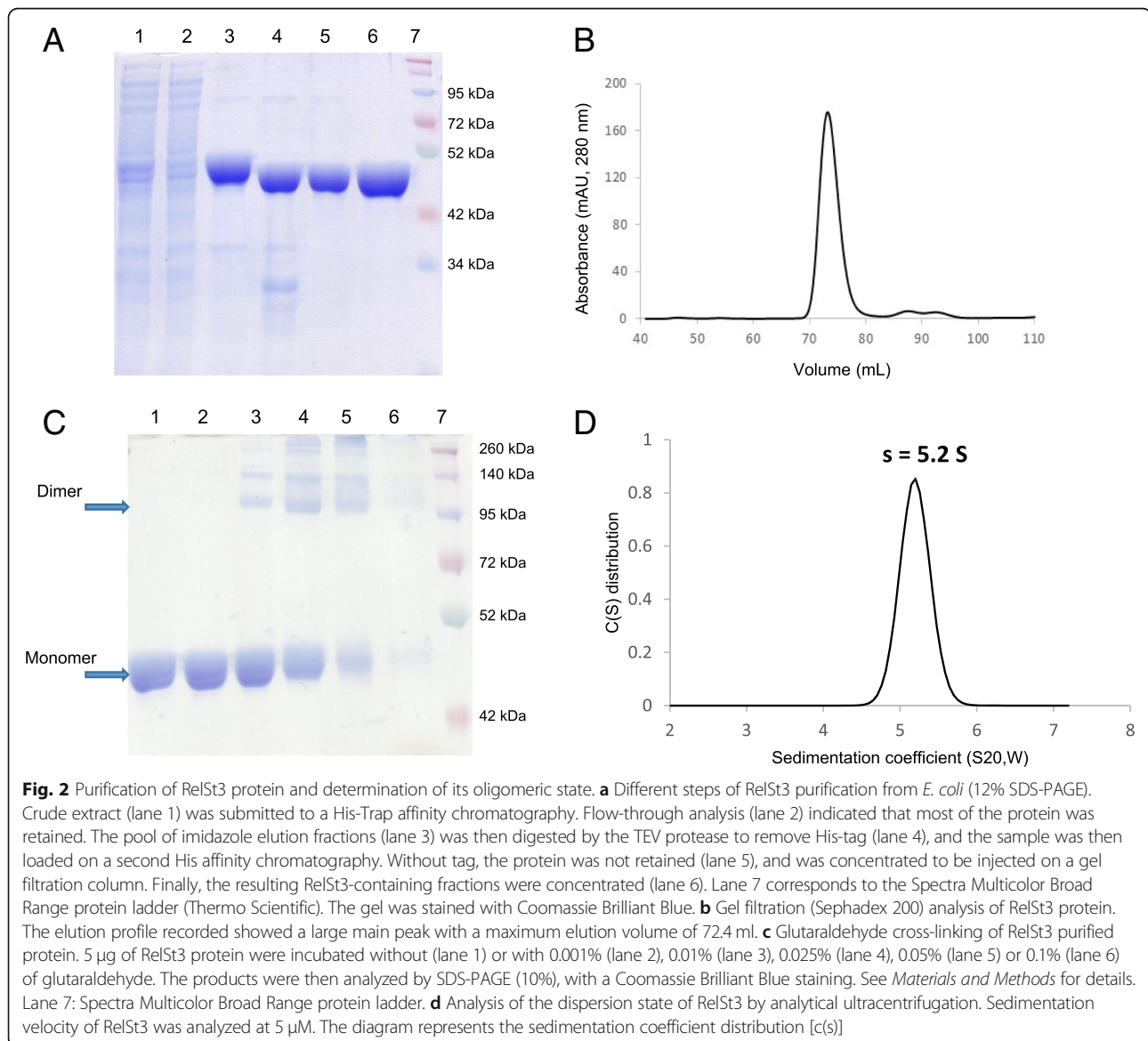
Donor strain	Recipient strain	Plasmid	ICES <sub>3</sub> transconjugants frequency	Plasmid mobilization frequency
LMG18311 (ICES <sub>3</sub> cat)	LMG18311 (pMG36e)	pOri1180	$8,8 \times 10^{-5}$	$< 10^{-6}$
LMG18311 (ICES <sub>3</sub> cat)	LMG18311 (pMG36e)	pOri1180-oriT	$2.0 \times 10^{-4}$	$2.5 \times 10^{-3}$

similar results. Most of the RelSt3 absorbance corresponded to a single peak, indicating the absence of higher aggregates, and demonstrating that the oligomeric form of RelSt3 is homogeneous at these protein concentrations. The Sedfit program was used to analyze the *c(s)* distribution model. The unique peak was associated with a

sedimentation coefficient (*S*<sub>20, w</sub>) of 5.2 *S*, a frictional ratio of 1.5 and a hydrodynamic radius of 4.4 nm. Taken together, these data indicate that the RelSt3 protein is an elongated dimer in solution.

### Secondary structure content and stability of RelSt3

Using various computational methods, the secondary structure of RelSt3 protein was predicted to consist of 41.7%  $\alpha$ -helices, 14.8%  $\beta$ -strands and 43.6% random coils (Table 2). Circular dichroism (CD) was used to analyze RelSt3 folding experimentally. Far-UV spectra were recorded at 20 °C. The spectral data were converted into mean residue ellipticity (Fig. 3a) and showed a maximum at 195 nm, a characteristic feature of  $\beta$ -strand structures, and two minima at 208 and 222 nm, typical of  $\alpha$ -helix content [53]. Deconvolution using



**Table 2** Secondary structure content of RelSt3

Prediction programs	$\alpha$ -helix	$\beta$ -strand	Turns	Unordered
PSI-pred	43.6	10.7	ND	45.7
SABLE	47.3	12.7	ND	40
Jpred4	42.2	14.4	ND	43.4
PredictProtein	38.5	15.1	ND	46.4
APSSP2	36.3	18.8	ND	44.9
Phyre2	42	17	ND	41
Average	41.7 $\pm$ 3.9	14.8 $\pm$ 2.9		43.6 $\pm$ 2.6
Deconvolution methods	$\alpha$ -helix	$\beta$ -strand	Turns	Unordered
Selcon3	39.7	21.4	17.7	21.8
ContinLL	35.3	26.7	18.4	19.6
CDSSTR	41.0	19.0	18.0	21.0
CDNN	43.2	10.4	14.0	35.3
Average	39.8 $\pm$ 3.3	19.4 $\pm$ 6.8	17.0 $\pm$ 2.0	24.4 $\pm$ 7.3

Data and standard deviation are given in percentages. ND not determined

different methods provided an average content of 39.8%  $\alpha$ -helix, 19.4%  $\beta$ -strands, 17.0% turns, and 24.4% unordered (Table 2), indicating that this experimentally determined secondary structure content is in agreement with the in silico prediction. This mixed  $\alpha/\beta$  proportion is similar to those of other CD-characterized relaxase proteins, including MobM from pMV158 and TraA from pIP501 [54, 55], as well as known structures of other relaxases [35, 36, 56, 57].

To estimate the stability of RelSt3 folding, thermal denaturation was monitored from 20 °C to 80 °C at 222 nm (typical minimum ellipticity of  $\alpha$ -helix) and at 195 nm (typical maximum ellipticity of  $\beta$ -strands). Approximately 75% of the ellipticity at 195 nm and 55% of the ellipticity at 222 nm were lost between 20 and 80 °C, with a one-step cooperative transition midpoint at 55 °C (Fig. 3b). After cooling back to 20 °C, almost no ellipticity was recovered (data not shown), indicating that the thermal denaturation is irreversible in our conditions. In conclusion, RelSt3 is a well-folded protein, but is sensitive to denaturation, that is consistent with the observation of protein aggregates during RelSt3 purification trials in presence of ionic force lower than 500 mM NaCl.

Using DSC experiments, we next investigated the overall stability of RelSt3. RelSt3 denaturation occurred in a single step with an enthalpy variation of 144 kcal/mol and a determined  $T_m$  of 50.95 °C (Fig. 3c), which is in good accordance with our CD results.

### RelSt3 structure model

To obtain further structural data of a MOB<sub>T</sub> relaxase, we built a 3D model of the tertiary structure of the RelSt3 relaxase domain (PF02486) by homology. As phylogenetic analysis indicated that RepSTK1 was the closest protein to

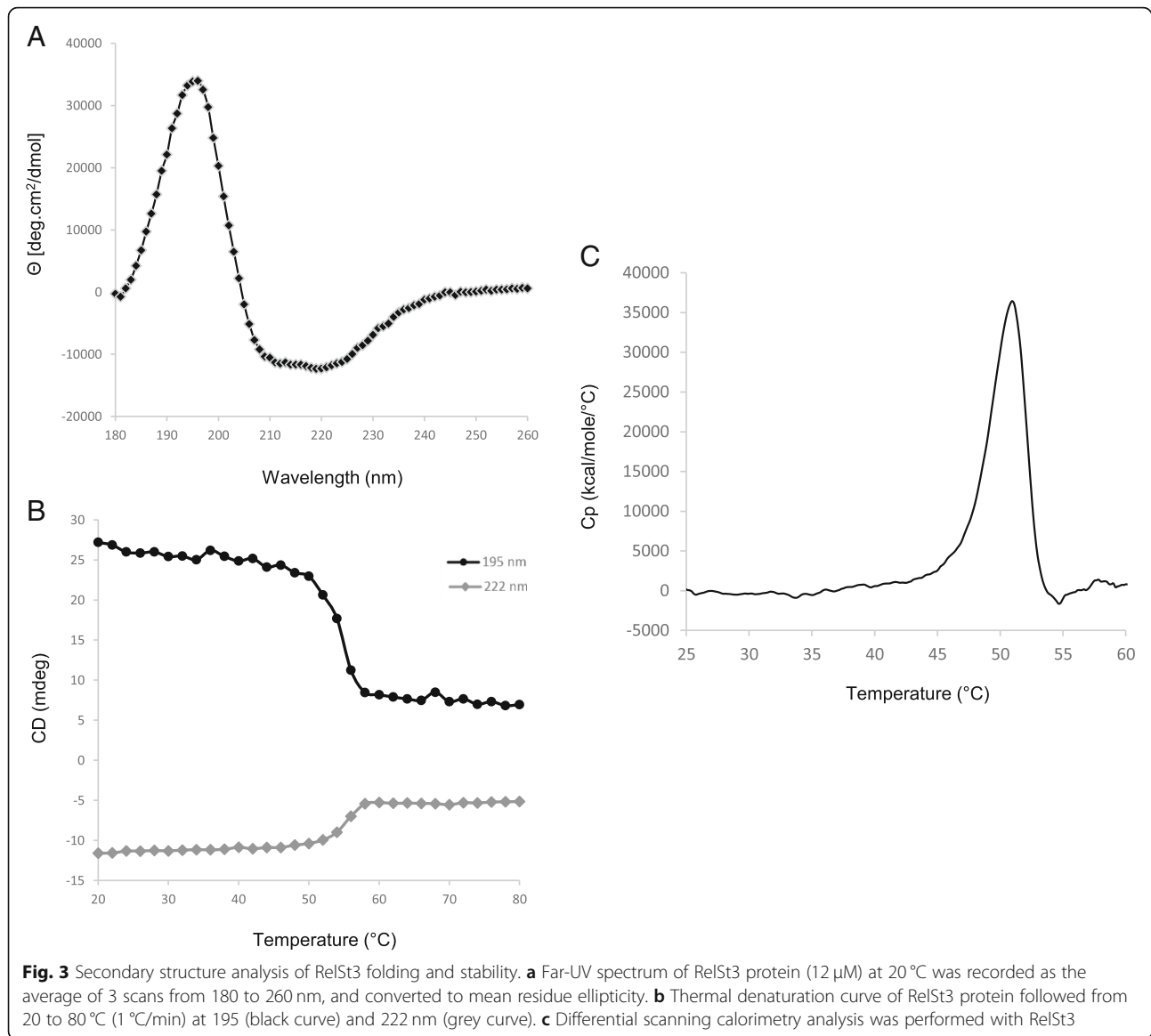
RelSt3 for which an experimental structure has been solved (Fig. 1), the X-ray structure of RepSTK1 (PDB 4CIJ) was used as a template. Like the RepSTK1 structure, the RelSt3 model is crescent-shaped, with two lobes of similar secondary structure patterns, each involving 5 beta-strands:  $\beta$ - $\alpha$ - $\alpha$ - $\beta$ - $\beta$ - $\beta$ - $\alpha$ - $\alpha$  (N-lobe) and  $\beta$ - $\alpha$ - $\alpha$ - $\beta$ - $\beta$ - $\alpha$ - $\alpha$ - $\beta$  (C-lobe) (Fig. 4) [34]. The underlined elements are absent from the template structure RepSTK1. The two 5-strand beta-sheets are adjacent and form a continuous 10-strand sheet, whose structure and position in the sequence are highly conserved with *Rep\_trans* proteins (despite RelSt3 and *Rep\_trans* proteins being poorly conserved in sequence, Fig. 1b). In our homology model of RelSt3 relaxase domain, the secondary and tertiary structures are very well conserved compared to the template, and stable over the dynamic simulation. The main differences are a shorter  $\beta$ 1- $\alpha$ 1 loop (20 vs 27 residues), a longer  $\beta$ 9- $\beta$ 10 linker (23 vs 12 residues) including an additional helix ( $\alpha$ 8), and shorter  $\beta$ 7 and  $\beta$ 8 strands (3–4 vs 8–9 residues). The parts of our model that correspond to the active site in the template are located at the junction of the two 5-strands sheets, involving the adjacent strands  $\beta$ 1 (N-lobe) and  $\beta$ 6- $\beta$ 9- $\beta$ 10 (C-lobe), and a short helix downstream of  $\beta$ 9 ( $\alpha$ 7, Figs. 4 and 7).

### In vitro relaxase activity of RelSt3

The relaxation activity of RelSt3 was assayed as previously described by Lorenzo-Díaz et al [54], using a pBR322 plasmid harbouring *ICEst3 oriT* (Additional file 4, Table S2). We were able to demonstrate a nicking activity on this supercoiled substrate (FI), leading to the formation of a relaxed form of the plasmid (FII) (Fig. 5a). To determine the optimal conditions for this relaxation activity, different temperatures (Fig. 5a), incubation times (Fig. 5b), and cations (Fig. 6a) were tested. As shown in Fig. 5a and b, a 20 min incubation at 30 °C of RelSt3 with supercoiled plasmid appeared to give a maximum amount of relaxed form FII. A small amount of linear plasmid was also observed (FIII), probably as a by-product from already nicked DNA. A dose-response experiment was also carried out (Fig. 5c) indicating that the nicking activity is dependent on RelSt3 concentration. As shown for *Rep\_trans* proteins [34, 58], no activity was observed without any cationic cofactor (Fig. 6a) indicating that RelSt3 activity depends on the presence of a cationic cofactor. Whereas  $\text{Ca}^{2+}$ ,  $\text{Zn}^{2+}$  and  $\text{Ni}^{2+}$  were not bona fide cofactors for RelSt3, the presence of  $\text{Co}^{2+}$ , and especially  $\text{Mg}^{2+}$  or  $\text{Mn}^{2+}$  is needed for RelSt3 relaxation activity (Fig. 6a). The RelSt3 activity is dependent on  $\text{Mn}^{2+}$  concentration, and a maximal conversion of form I was observed above 1 mM  $\text{Mn}^{2+}$  (Fig. 6b).

### Characterization of RelSt3 variants affected in their nicking activity

As mentioned above, multiple alignment of protein sequences (Additional files 1 and 2: Figure S1 and S2) together

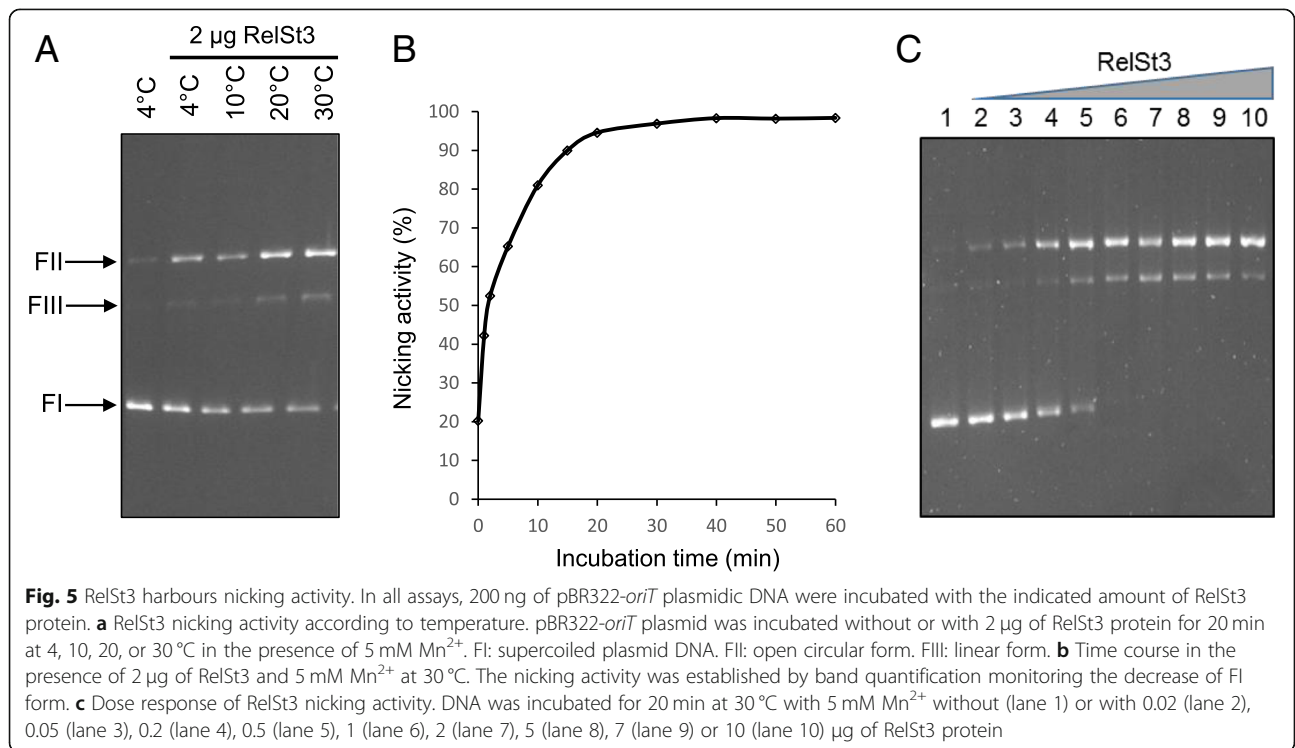
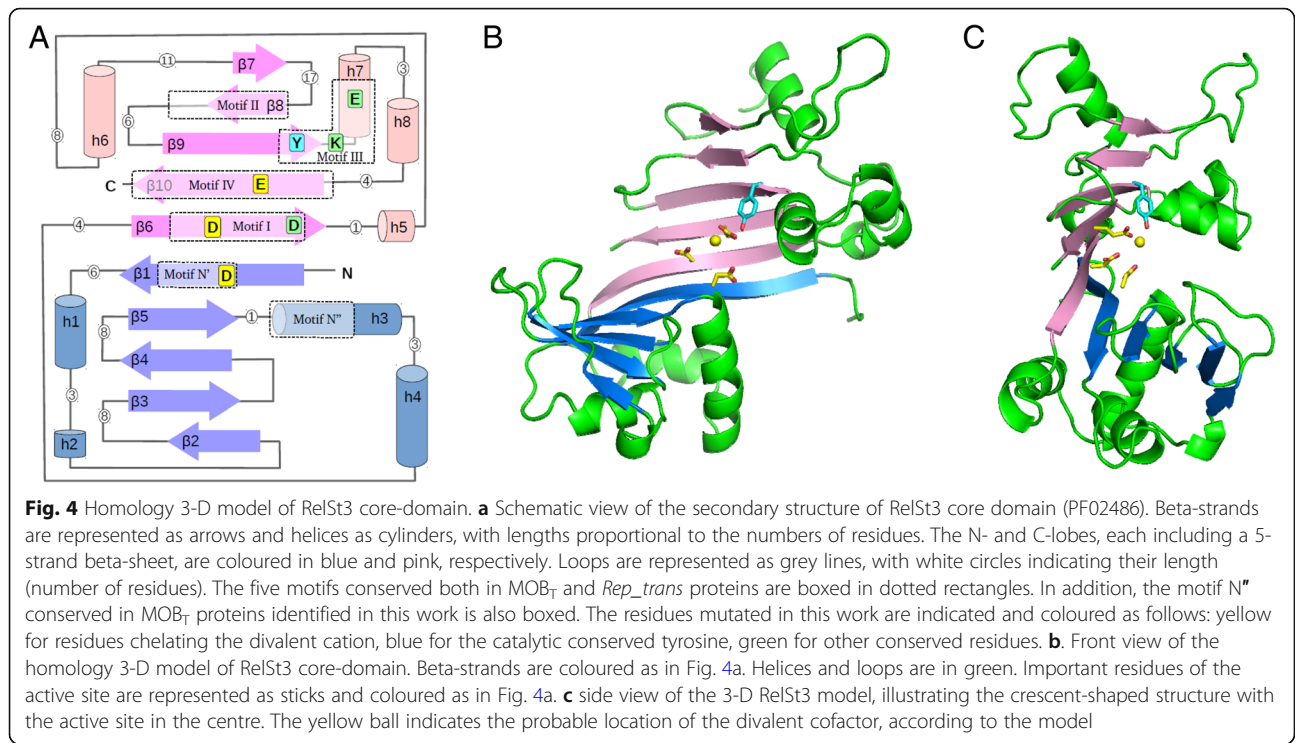


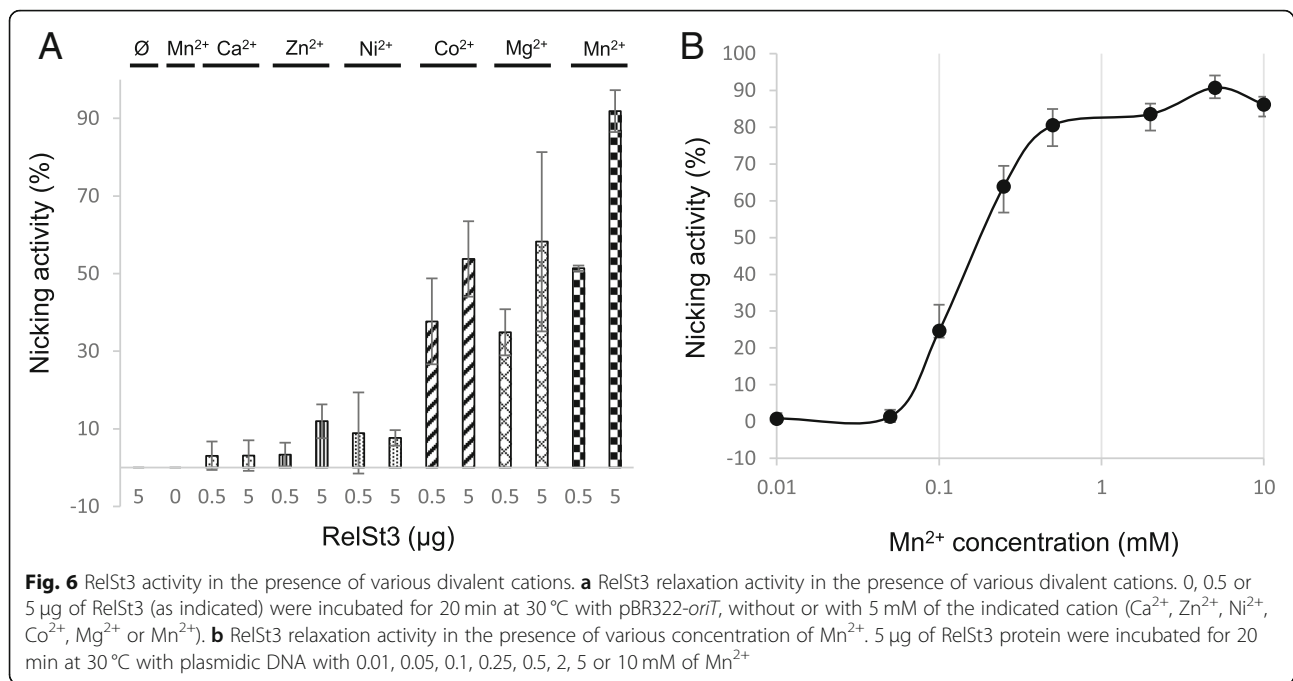
with our 3D homology model (Fig. 4) identified motifs in RelSt3 similar to conserved motifs proposed for *Rep\_trans* proteins [34]. These motifs are also found in other MOB<sub>T</sub> relaxases (Fig. 1c). Carr et al (2016) demonstrated that 3 acidic residues located in 3 different motifs (N', I and IV, shown by a star in Additional file 1: Figure S1) are involved in the coordination of the cationic cofactor in *Rep\_trans* proteins. As these acidic residues are conserved in the RelSt3 sequence (D72, D177, E280, Fig. 1c and Additional file 1: Figure S1), each one was independently mutated into alanine in order to assess its involvement in RelSt3 activity. In motif III (Additional file 1: Figure S1), the RelSt3 tyrosine Y252 is conserved within *Rep\_trans* proteins from plasmids of the pT181 family. The corresponding residue (Y188) of the RepD protein from the staphylococcal plasmid pC221 has already been demonstrated to be catalytic [49]. To confirm

that this residue is also catalytic in a MOB<sub>T</sub> protein, a Y252A variant of RelSt3 was generated. Two other conserved residues were also mutated, K254 and E257, located in the vicinity of the putative catalytic Y252 residue in motif III. All variants were expressed using the same expression vector and with the same purification protocol as for the purification of the WT protein. The nicking activity of each variant was compared to the WT (Fig. 5c and Fig. 7a) with a standard dose-response experiment using Mn<sup>2+</sup> as cofactor.

As shown in Fig. 7a, each RelSt3 single mutant D72A, D177A and E280A had completely lost the nicking activity, which is consistent with the results obtained for *Rep\_trans* proteins modified at these conserved positions [34]. The Y252A variant also showed an absence of nicking activity, suggesting that this residue plays a role of catalytic tyrosine for RelSt3. Concerning the two other conserved residues in







motif III, K254 appeared to be essential, whereas E257 was found to be important but not essential for DNA relaxation (Fig. 7a). The positions of these residues in the active site structure model of RelSt3 (Fig. 7b) suggest that E257 could be involved in maintaining the catalytic Tyr in the correct position, whereas K254 could be involved in the stabilization of the acidic residues assumed to coordinate the cation (especially D72 and E280, with Calpha-Calpha distances of 10.5 and 13.0 Å respectively, and maximal side-chain sizes of 6.4 Å, 3.8 Å and 4.8 Å for K254, D72 and E280, respectively).

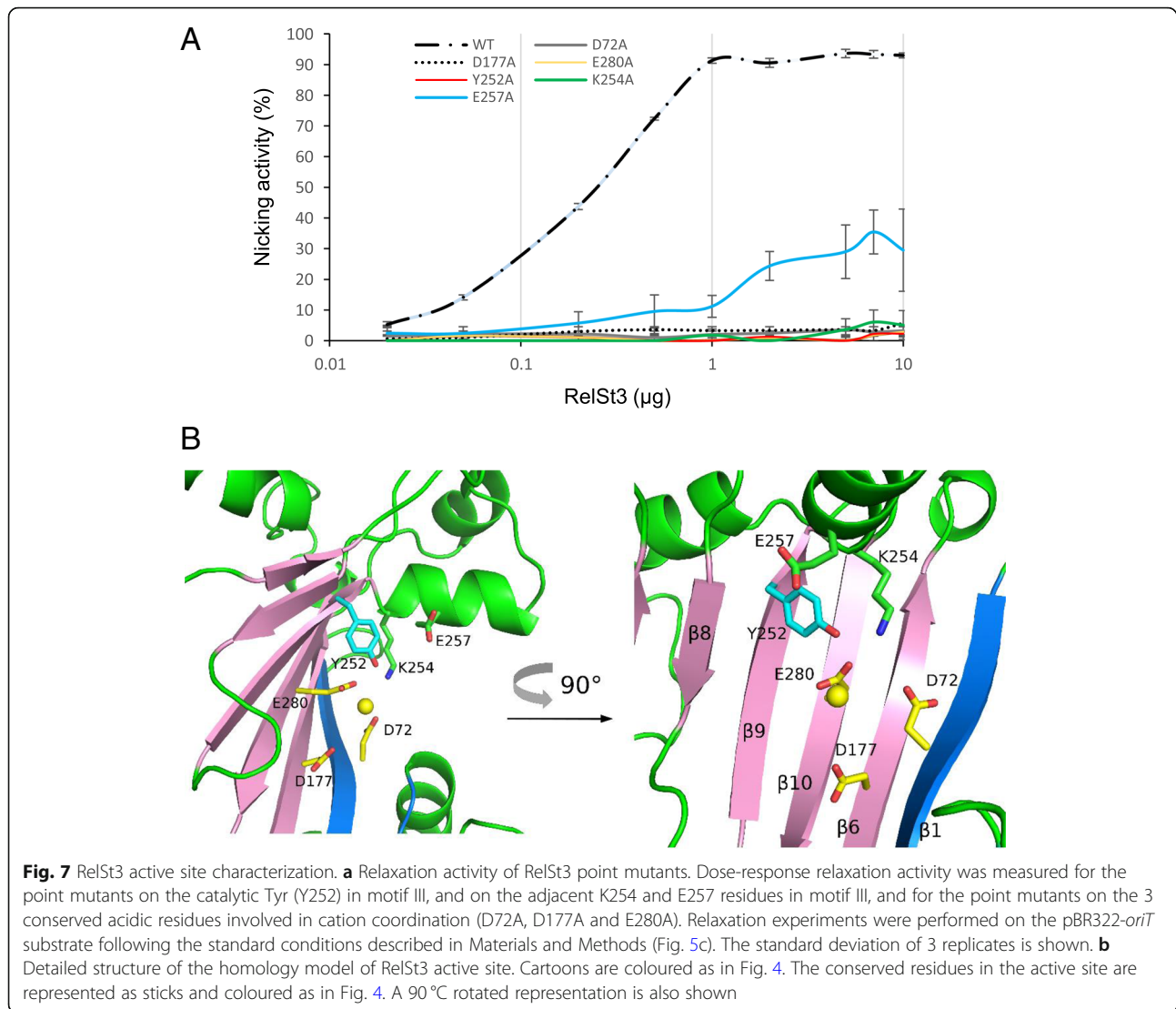
## Discussion

In this work, we undertook the biochemical characterization of the RelSt3 protein encoded by *ICES3* of *Streptococcus thermophilus*, a MOB<sub>T</sub> relaxase. MOB<sub>T</sub> relaxases are encoded by ICEs belonging to the *ICES3*/*ICEBs1*/*Tn916* superfamily found in many Firmicutes [15]. We highlighted here the relationship between the MOB<sub>T</sub> family of relaxases and the RCR initiator proteins of the *Rep\_trans* family, including Rep proteins encoded by the pT181 plasmid family and the structurally characterized RepSTK1 from *Geobacillus stearothermophilus* [34]. Interestingly, an activity of RCR initiation was discovered for several MOB<sub>T</sub> relaxases encoded by ICEs (Orf20 and Nick). Indeed, these relaxases were shown to be responsible for autonomous RCR initiation starting from the *oriT* [32, 59]. Reciprocally, it has already been demonstrated that RCR initiators involved in the maintenance of small plasmids from Firmicutes can also act as conjugative relaxases. Indeed, these plasmids are still mobilized *in trans* by

*ICEBs1* deleted from its relaxase gene [28]. However, this mobilization was only demonstrated for RCR plasmids encoding a *Rep\_1* initiator protein, and was not tested for a RCR plasmid encoding a *Rep\_trans* protein.

The purified RelSt3 protein was demonstrated to be homogeneously dimeric. This dimeric state is in agreement with the *Rep\_trans* 3D structures solved recently, where the corresponding proteins crystallized as dimers [34]. We demonstrated the relaxation activity of RelSt3 on a plasmidic DNA harboring *ICES3 oriT*, confirming the functionality of this protein as a relaxase for *ICES3*. As observed for *Rep\_trans* proteins, this single-strand endonuclease activity is dependent on the presence of a divalent metal cofactor that turned out to be preferentially Mg<sup>2+</sup> or Mn<sup>2+</sup>. Although historically the activity of most relaxases was mainly tested with Mg<sup>2+</sup> [22, 60–62], when compared with Mn<sup>2+</sup> the latter was found (i) to be more efficient for nicking activity [54, 63], or (ii) to have a better affinity as measured by micro-calorimetry [54, 64], or (iii) to be the most probable cation according to structural and ICP-MS data [35]. We also noticed that Mn<sup>2+</sup> was identified in the RepSTK1 structure provided by Carr et al [34]. Accordingly, we observed a slightly higher efficiency in relaxation assays for RelSt3 with Mn<sup>2+</sup> compared to Mg<sup>2+</sup>.

The secondary structure of the RelSt3 protein deduced from the circular dichroism analysis is very similar to the *in silico* predictions. A mix of α-helices and β-strands is observed as for most, if not all, characterized RCR Rep and Mob proteins [21, 26, 34, 39]. The thermal stability of a few canonical relaxases has already been studied using circular dichroism, and the addition of a



cationic cofactor induced a stabilization of the protein at higher temperatures. Whereas the MobA\_R1162 and TraA\_pIP501 (MOB<sub>Q</sub>) relaxases present a unique and short transition with a midpoint around 42 °C, indicating a cooperative unfolding [55, 63], the MobM\_pMV158 (MOB<sub>V</sub>) relaxase harbors a complex denaturation profile with two transitions (midpoints at 21 and 42 °C) spanning more than 30 °C [54]. In comparison, the MOB<sub>T</sub> relaxase RelSt3 exhibits a one-step transition as shown by CD and DSC experiments, with midpoints at 55 and 51 °C, respectively. This is clearly higher than observed for other characterized relaxases, indicating that RelSt3 could be intrinsically more stable.

Phenotype analyses of RelSt3 point mutations confirmed the relationship between MOB<sub>T</sub> and *Rep\_trans* proteins. We showed that the catalytic tyrosine of RelSt3 is located in the conserved motif III as in *Rep\_trans* proteins. This is the first characterization of catalytic motif

for a MOB<sub>T</sub> relaxase. We also demonstrated the involvement of the 3 conserved acidic residues D72, D177 and E280 (located in motifs N', I and IV, respectively) in relaxation activity, as demonstrated for RepD variants [34]. These results suggest that both MOB<sub>T</sub> and *Rep\_trans* families (i.e. the PF02486 domain) use 3 distant acidic residues in primary sequence to coordinate their cationic cofactor. In addition, we analyzed the involvement of two other conserved residues of motif III in RelSt3 activity, K254 and E257. Both residues are important for RelSt3 nicking activity, indicating that they could assume a structural role within the active site. Indeed, our 3D structural model of the PF02486 RelSt3 domain suggests that E257 could interact with the catalytic Tyr, and that K254 could interact with the acidic residues D72 and E280 that are assumed to coordinate the cation, maintaining them in the correct position.

The 3D structure model based on RepSTK1 crystal structure revealed a fold that resembles classical HxH endonucleases [21], as it presents a central 10-strand beta sheet (2 × 5 strands) decorated by several alpha-helices. Whereas HxH endonucleases harbour their 3 histidine residues that coordinate the cationic cofactor on two of the central beta-sheets, RelSt3 as well as *Rep\_trans* structures instead use 3 acidic residues to do so, each located on a different strand of the central sheet. This correlates with the quite important distance between each of these residues in the primary sequence, lying respectively in motifs N', I and IV, compared to the classical 3 His that are close to each other in the primary sequence. Another striking difference with HxH endonucleases is that RelSt3 and *Rep\_trans* structures harbour their conserved catalytic Tyr residue on one of the central beta-strands (strand beta9 in RelSt3 model), whereas the HxH proteins have their catalytic Tyr residue(s) on an alpha-helix close to the cationic cofactor. These observations suggest that the catalytic assembly orchestrating a Tyr residue and a divalent cation to cleave a DNA substrate arose at least twice in evolution, giving rise to the HxH endonucleases on one hand, and to the *Rep\_trans*/MOB<sub>T</sub> families (harbouring the PF02486 domain) on the other hand, as a remarkable example of evolutionary convergence.

## Conclusions

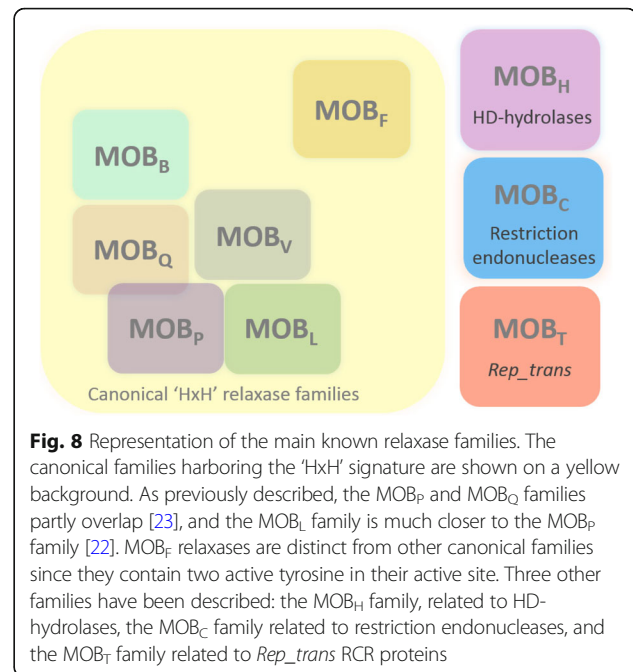
Even if two MOB<sub>T</sub> relaxases were identified circa 10 years ago (for Tn916 and ICEBs1 conjugative elements) [31, 33], their affiliation to a unique relaxase family was only pointed out later. This family was unclassified in 2009 [23], proposed in 2011 [7], but surprisingly, it was not mentioned recently [22, 39]. Taken together, our data commit us to propose the MOB<sub>T</sub> family as a full non-canonical relaxase family alongside the two other previously described non canonical families (MOB<sub>H</sub> and MOB<sub>C</sub>) (Fig. 8).

In conclusion, we performed the first biochemical characterization of a member of the MOB<sub>T</sub> family of relaxases, the RelSt3 protein. Its active site characterized here constitutes a potential target for the design of conjugation inhibitors. With this aim in mind, further work is needed, in particular to determine an experimental 3D structure of a MOB<sub>T</sub> protein.

## Methods

### Phylogenetic analysis and identification of conserved motifs

BLAST searches were performed using the NCBI non redundant (nr) database. The multiple alignment for the phylogenetic analysis was built with the MultAlin interface [65]. Only homologous positions (54 in total) were selected manually for phylogenetic analysis. A tree was constructed with PhyML 3.1 (downloaded from [http://](http://www.atgc-montpellier.fr/phyml/versions.php)



**Fig. 8** Representation of the main known relaxase families. The canonical families harboring the 'HxH' signature are shown on a yellow background. As previously described, the MOB<sub>P</sub> and MOB<sub>Q</sub> families partly overlap [23], and the MOB<sub>L</sub> family is much closer to the MOB<sub>P</sub> family [22]. MOB<sub>F</sub> relaxases are distinct from other canonical families since they contain two active tyrosine in their active site. Three other families have been described: the MOB<sub>H</sub> family, related to HD-hydrolases, the MOB<sub>C</sub> family related to restriction endonucleases, and the MOB<sub>T</sub> family related to *Rep\_trans* RCR proteins

[www.atgc-montpellier.fr/phyml/versions.php](http://www.atgc-montpellier.fr/phyml/versions.php)) [66]. The JTT model of amino acid substitution was chosen [67], and a gamma correction with four classes of sites was used. The alpha parameter and the proportion of invariable sites were estimated. The robustness of the tree was tested with PhyML by non-parametric bootstrap analysis (100 replicates).

To identify the conserved motifs in MOB<sub>T</sub> relaxases, we first constructed a set of MOB<sub>T</sub> sequences using NCBI's CDART (Conserved Domain Architecture Retrieval Tool) application, with the RelSt3 sequence as a query [68]. All sequences harboring the RelSt3 architecture (N-terminal HTH domain and PF02486 domain) were aggregated together with the sequences harboring the Nick domain (PF02486 domain only, lacking the HTH domain). A total of 2016 sequences were then clustered using the CD-HIT suite [69], giving rise to 711, 383 and 261 clusters according to sequence identity cut-off of 98, 95 and 90%, respectively. One member of each cluster was selected for sequence alignment with Muscle [70] or ClustalW [71], and the resulting alignments were submitted to the WebLogo 3 interface [72]. The resulting files were very similar for all sequence identity cut-offs (used for CD-HIT clustering) and multiple alignment programs used.

### Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Additional file 4: Tables S1 and S2. *Streptococcus thermophilus* LMG18311 harboring ICESt3 labelled with a chloramphenicol resistance gene [51], or its derivative depleted of the *orf* gene, were used as donor strains during mating experiments. A strain that derives from *S.*



*thermophilus* LMG18311 and harbors a pMG36e plasmid carrying erythromycin resistance was used as recipient cell. *E. coli* EC101 was used as the host for cloning procedures with pG<sup>+</sup>host9 [73]. For heterologous protein production in *E. coli*, constructs were cloned into the DH5 $\alpha$  strain, and over-expression was performed in BL21(DE3) cells.

#### Construction of ICES $t3\Delta orfJcat$ and of a mobilizable *oriT*-harboring plasmid

The deletion strategy used to construct the *S. thermophilus* ICES $t3\Delta orfJcat$  strain is derived from Biswas et al, 1993 [74], with several adaptations described in Additional file 4, Supplementary Materials and Methods. Three independent mutants were verified by sequencing and then used in conjugation assays.

The mobilizable *oriT*-harboring plasmid was constructed using the pOri23 backbone [75], where the Erm<sup>R</sup> cassette was replaced by a Spec<sup>R</sup> cassette, giving rise to the pOri1180 plasmid. In order to experimentally identify the *oriT* region of ICES $t3$ , we constructed a plasmid encompassing the intergenic region between *orfK* and *orfJ* using *EcoRI* and *ApaI* restriction sites (according to primers described in Additional file 4, Table S3), giving rise to the pOri1180-*oriT* plasmid.

#### Mating experiments

Mating experiments were done as described in Dahmane et al, 2017 [76]. Briefly, donor and recipient strains were grown overnight with appropriate antibiotics. Cultures diluted to 1:100 were grown without any antibiotic until mid-exponential phase, and then mixed and concentrated 30 times in LM17 broth. For pOri1180 mobilization experiments, the donor strains were grown in the presence of spectinomycin (500  $\mu\text{g}/\text{mL}$ ) until mid-exponential phase and washed before the mating experiment. For each mating experiment, three 150  $\mu\text{L}$  aliquots of cells were spread separately on 0.45  $\mu\text{m}$  pore-size nitrocellulose filters (Millipore) deposited on soft agar (0.8%) LM17 plates and then incubated overnight at 42  $^{\circ}\text{C}$ . To select transconjugant cells, bacteria recovered from the filters were directly spread or concentrated 10 times before plating on chloramphenicol and erythromycin, or spectinomycin and erythromycin containing-LM17. After dilutions, the donor and the recipient cells were selected on LM17 plates containing the appropriate antibiotics. After a 24 h-incubation at 42  $^{\circ}\text{C}$ , CFUs counting of three independent biological replicates enabled us to quantify the mating frequency as the number of transconjugants per donor cell.

#### Cloning of *orfJ* gene for production in *E. coli* and mutagenesis

Mutagenesis was performed by overlap PCR with the oligonucleotides described in Additional file 4, Table S3.

To produce RelSt3 WT or RelSt3 variants, the *orfJ* gene (*wt* or mutated) was cloned in *E. coli* DH5 $\alpha$ , using *NdeI* and *HindIII* restriction sites into pSKB3 vector (derived from pET28a plasmid, gift from Stephen K. Burley), in frame with a 6 His-tag and a TEV protease cleavage site (N-terminal). Over-expression and purification of recombinant RelSt3 proteins are described in Additional file 4, Supplementary Materials and Methods.

#### Protein cross-linking

RelSt3 (5  $\mu\text{g}$ ) was incubated in a final reaction volume of 20  $\mu\text{L}$  with various concentrations of glutaraldehyde (0.0, 0.001, 0.01, 0.025, 0.05 and 0.1% v/v). The reaction was stopped after 15 min incubation at room temperature by addition of 1 M glycine pH 8.0 to a final concentration of 90 mM. After a 5 min incubation at room temperature, Laemmli loading buffer was added and the samples were heated at 100  $^{\circ}\text{C}$  for 5 min prior to load on a 10% SDS-polyacrylamide gel.

#### Analytical ultracentrifugation analysis

Sedimentation velocity assays with RelSt3 were performed at 1 and 5  $\mu\text{M}$  in buffer 20 mM Tris-HCl pH 7.5, 500 mM NaCl. Samples were centrifuged for 16 h at a rotor speed of 42,000 rpm in a Proteomelab XL-I analytical ultracentrifuge (Beckman Coulter) at 25  $^{\circ}\text{C}$  in a four hole rotor AN60-Ti equipped with 12 mm double-sector cells with epon centrepieces. Detection of the protein concentration as a function of radial position and time was performed by absorbance measurements at a wavelength of 280 nm. The following parameters were calculated using Sednterp 1.09 (<http://www.jphilo.mailway.com/download.htm>) and used for the analysis of the experiment: partial specific volume  $\bar{v} = 0.734$  mL/g, viscosity  $\eta = 0.01054$  P, and density  $\rho = 1.0192$  g/mL. Sedimentation velocity data analysis was performed by continuous size distribution analysis *c(s)* using the program Sedfit 15.1 (<http://www.analyticalultracentrifugation.com>). All the *c(s)* distributions were calculated with a fitted frictional ratio  $f/f_0$  and a maximum entropy regularization procedure with a confidence level of 0.95.

#### Secondary structure prediction, circular dichroism (CD) and differential scanning calorimetry (DSC) assays

Secondary structure content was predicted using the PSI-pred [77], SABLE [78], JPred4 [79], PredictProtein [80], APSSP [81] and Phyre2 [47] programs.

Far-UV CD spectra were obtained from 180 to 260 nm with a Chirascan Plus spectropolarimeter (Applied Photophysics, Ltd., UK) equipped with a Peltier temperature control unit (20  $^{\circ}\text{C}$ ). RelSt3 protein was loaded at 20  $\mu\text{M}$  in CD buffer (50 mM sodium phosphate pH 7.0, 500 mM NaF) on a flat quartz cell of 0.1 mm path length. Spectra



were recorded with a scan speed of 60 nm.min<sup>-1</sup> and 1 nm spectral bandwidth. After an average of 3 scans and subtraction of buffer contribution, raw data were converted to mean residue ellipticity (deg.cm<sup>2</sup>.dmmol<sup>-1</sup> per residue) using Pro-Data Viewer software (Applied Photo-physics, Ltd., UK), and was deconvolved using SELCON3, CONTINLL and CDSSTR algorithms available at the DichroWeb site [82, 83], with the reference data set #3. Data were also deconvolved with the CDNN program [84]. Thermal denaturation curves were followed at 195 and 222 nm from 20 to 80 °C (1 °C/min), and finally a 180–260 nm spectrum was recorded after cooling the sample back to 20 °C.

Differential scanning calorimetry (DSC) experiments were performed with a VP-DSC microcalorimeter (Microcal, Malvern Panalytical, Vénissieux, France). 500 µL of 6.6 µM RelSt3 was incubated with a heating rate of 60 °C/h in buffer A (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 2 mM β-mercapto-ethanol). The experiments were done twice.

### Homology modeling

Ten 3D models of the tertiary structure of the RelSt3 PF02486 domain (residues C61–D285) were built by homology with MODELLER [85], using as a template the X-ray structure of the RepSTK1 protein (PDB 4CIJ, 30% sequence identity with the RelSt3 PF02486 domain). The model with the best Discrete Optimized Protein Energy (DOPE) score [86] was then refined by a molecular dynamics simulation of 100 ns in explicit solvent, and the most representative structure was selected by clustering the conformations selected from each ps of the simulation. Molecular dynamics simulation were performed with the NAMD software [87], using the charmm36 force field [88], and a box of explicit TIP3P water molecules. Counter-ions were added to neutralize the system.

### Supercoiled plasmidic DNA relaxation assays

Relaxation assays were adapted from Lorenzo-Diaz et al, 2011 [54]. Plasmidic DNA (200 ng) of pBR322-*oriT* (Additional file 4, Table S2) was incubated 20 min at 30 °C with increasing amounts of RelSt3 protein (WT or variants, see figure legends), in R buffer [20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM MnCl<sub>2</sub>, 0.05 mM EDTA, 1% glycerol (*w/v*) and 1 mM β-mercapto-ethanol] with a final volume of 20 µL. Reactions were stopped by addition of final concentrations of 0.5% SDS and 100 µg/mL of proteinase K, followed by a second incubation at 30 °C for 20 min. Samples were electrophoresed on 1% agarose gel in Tris-Borate-EDTA-1X for 20 h at 2.5 V/cm and the gel was stained with ethidium bromide. Gel pictures were captured using a Bio-Rad XR+ system, and

bands were quantified with Image Lab software (Bio-Rad Laboratories). Variations from these standard conditions are indicated in the figure legends.

### Additional files

**Additional file 1: Figure S1.** Alignment of MOB<sub>T</sub> relaxase representatives with *Rep\_trans* RCR initiator representatives. This alignment with reduced number of sequences allows the identification of the different motifs cited in this study. See complete legend in Additional file 4. (RTF 194 kb)

**Additional file 2: Figure S2.** Sequence alignment of MOB<sub>T</sub> relaxase and *Rep\_trans* RCR initiators. This larger alignment with more MOB<sub>T</sub> and *Rep\_trans* initiator sequences illustrates the conservation of the motifs through very distant proteins. See complete legend in Additional file 4. (RTF 1402 kb)

**Additional file 3: Figure S3.** Calibration curve of the gel filtration column. A Sephadex S200 HiLoad 16/60 column (GE Healthcare) was used as a final step of purification of RelSt3 protein. This column was calibrated and allowed us to estimate the apparent molecular weight of RelSt3. See complete legend in Additional file 4. (PPTX 40 kb)

**Additional file 4:** Supplementary material text. This text includes supplementary Materials and Methods, complete Additional file legends, and tables for bacterial strains, plasmids and oligonucleotides used in this study. (DOC 201 kb)

### Abbreviations

CD: Circular dichroism; DSC: Differential scanning calorimetry; HGT: Horizontal gene transfer; ICE: Integrative and conjugative elements; *oriT*: origin-of-transfer; RCR: Rolling-circle replication; T4SS: Type-IV secretion system

### Acknowledgements

We thank Alexandre Kriznik from the Biophysics and Structural Biology Platform, part of UMS2008 IBSLor (Université de Lorraine, CNRS, INSERM), for his advices and technical support during CD and DSC experiments. We are particularly grateful to Bertrand Raynal from the molecular biophysics platform of the Institut Pasteur in Paris for analytical ultracentrifugation experiments. We also thank Rémi Leblay and Mégane Winterstein for technical assistance.

### Funding

This work was supported by the Université de Lorraine and the Région Grand-Est. This work was co-funded by the french PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE, and by the European Union through the Regional Operational Program of the European Regional Development Fund (ERDF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request. The accession number of all protein sequences used for sequence alignments or phylogenetic analysis are listed in the Additional file 4, Table S4.

### Authors' contributions

NS, SP, MDD and NLB designed experiments. ER, ICB, PM, VL, BM and JS performed experiments. NS, ICB, VL, BM, GG, SP, MDD and NLB analyzed the data. NS wrote the original draft of the manuscript. ICB, VL, DWR, GG, SP, MDD and NLB reviewed the manuscript. This work was supervised by NS, DWR, MDD and NLB. All the authors approved the final manuscript.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Université de Lorraine, Inra, UMR1128 DynAMic, F-54000 Nancy, France.

<sup>2</sup>Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France.

Received: 15 March 2019 Accepted: 11 April 2019

Published online: 03 May 2019

**References**

- Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol*. 2005;3:711–21. <https://doi.org/10.1038/nrmicro1234>.
- Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol*. 2005;3:722–32. <https://doi.org/10.1038/nrmicro1235>.
- Wozniak RAF, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol*. 2010;8:552–63. <https://doi.org/10.1038/nrmicro2382>.
- Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev*. 2014;38:720–60. <https://doi.org/10.1111/1574-6976.12058>.
- Burrus V, Pavlovic G, Decaris B, Guédon G. Conjugative transposons: the tip of the iceberg. *Mol Microbiol*. 2002;46:601–10.
- Roberts AP, Mullany P. Tn 916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev*. 2011;35:856–71. <https://doi.org/10.1111/j.1574-6976.2011.00283.x>.
- Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. The repertoire of ICE in prokaryotes underscores the Unity, diversity, and ubiquity of conjugation. Casadesús J, editor. *PLoS Genet*. 2011;7:e1002222. <https://doi.org/10.1371/journal.pgen.1002222>.
- Top J, Willemis R, Bonten M. Emergence of CC17 *Enterococcus faecium*: from commensal to hospital-adapted pathogen. *FEMS Immunol Med Microbiol*. 2008;52:297–308. <https://doi.org/10.1111/j.1574-695X.2008.00383.x>.
- Huang J, Chen L, Li D, Wang M, Du F, Gao Y, et al. Emergence of a vanG-carrying and multidrug resistant ICE in zoonotic pathogen *Streptococcus suis*. *Vet Microbiol*. 2018;222:109–13. <https://doi.org/10.1016/j.vetmic.2018.07.008>.
- Haenni M, Lupo A, Madec J-Y. Antimicrobial Resistance in *Streptococcus* spp. *Microbiol Spectr*. 2018;6:eARBA-0008-2017. <https://doi.org/10.1128/microbiolspec.ARBA-0008-2017>.
- Sansever EA, Robinson DA. Staphylococci on ICE: overlooked agents of horizontal gene transfer. *Mob Genet Elem*. 2017;7:1–10. <https://doi.org/10.1080/2159256X.2017.1368433>.
- Brouwer MSM, Warburton PJ, Roberts AP, Mullany P, Allan E. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. Highlander SK, editor. *PLoS One*. 2011;6:e23014. <https://doi.org/10.1371/journal.pone.0023014>.
- Adams V, Han X, Lyras D, Rood JI. Antibiotic resistance plasmids and mobile genetic elements of *Clostridium perfringens*. *Plasmid*. 2018:e2018.07.002. <https://doi.org/10.1016/j.plasmid.2018.07.002>.
- Santoro F, Vianna ME, Roberts AP. Variation on a theme; an overview of the Tn916/Tn1545 family of mobile genetic elements in the oral and nasopharyngeal streptococci. *Front Microbiol*. 2014;5:e0535. <https://doi.org/10.3389/fmicb.2014.00535>.
- Ambroset C, Coluzzi C, Guédon G, Devignes M-D, Loux V, Lacroix T, et al. New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front Microbiol*. 2016;6:e1483. <https://doi.org/10.3389/fmicb.2015.01483>.
- Schroeder MR, Stephens DS. Macrolide Resistance in *Streptococcus pneumoniae*. *Front Cell Infect Microbiol*. 2016;6:e98. <https://doi.org/10.3389/fcimb.2016.00098>.
- De La Cruz F, Frost LS, Meyer RJ, Zechner EL. Conjugative DNA metabolism in gram-negative bacteria. *FEMS Microbiol Rev*. 2010;34:18–40. <https://doi.org/10.1111/j.1574-6976.2009.00195.x>.
- Zechner EL, Moncalián G, de la Cruz F. Relaxases and plasmid transfer in gram-negative bacteria. In: Backert S, Grohmann E, editors. *Type IV secretion in gram-negative and gram-positive bacteria*. Cham: Springer International Publishing; 2017. p. 93–113. [https://doi.org/10.1007/978-3-319-75241-9\\_4](https://doi.org/10.1007/978-3-319-75241-9_4).
- Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I. Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev*. 2014;39:81–95. <https://doi.org/10.1111/1574-6976.12085>.
- Draper O, César CE, Machón C, de la Cruz F, Llosa M. Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proc Natl Acad Sci U S A*. 2005;102:16385–90.
- Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol*. 2013;11:525–38. <https://doi.org/10.1038/nrmicro3067>.
- Ramachandran G, Miguel-Arribas A, Abia D, Singh PK, Crespo I, Gago-Cordoba C, et al. Discovery of a new family of relaxases in Firmicutes bacteria. Hughes D, editor. *PLoS Genet*. 2017;13:e1006586. <https://doi.org/10.1371/journal.pgen.1006586>.
- Garcillán-Barcia MP, Francia MV, de la Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev*. 2009;33:657–87. <https://doi.org/10.1111/j.1574-6976.2009.00168.x>.
- Ilyina TV, Koonin EV. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res*. 1992;20:3279–85.
- Salgado-Pabon W, Jain S, Turner N, van der Does C, Dillard JP. A novel relaxase homologue is involved in chromosomal DNA processing for type IV secretion in *Neisseria gonorrhoeae*. *Mol Microbiol*. 2007;66:930–47. <https://doi.org/10.1111/j.1365-2958.2007.05966.x>.
- Francia MV, Clewell DB, de la Cruz F, Moncalian G. Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases. *Proc Natl Acad Sci*. 2013;110:13606–11. <https://doi.org/10.1073/pnas.1310037110>.
- Wisniewski JA, Traore DA, Bannam TL, Lyras D, Whistock JC, Rood JI. TcplM: a novel relaxase that mediates transfer of large conjugative plasmids from *Clostridium perfringens*: the novel relaxase TcplM. *Mol Microbiol*. 2016;99:884–96. <https://doi.org/10.1111/mmi.13270>.
- Lee CA, Thomas J, Grossman AD. The *Bacillus subtilis* conjugative transposon ICEBs1 mobilizes plasmids lacking dedicated mobilization functions. *J Bacteriol*. 2012;194:3165–72. <https://doi.org/10.1128/JB.00301-12>.
- Coluzzi C, Guédon G, Devignes M-D, Ambroset C, Loux V, Lacroix T, et al. A glimpse into the world of integrative and Mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. *Front Microbiol*. 2017;8. <https://doi.org/10.3389/fmicb.2017.00443>.
- Douarre P-E, Sauvage E, Poyart C, Glaser P. Host specificity in the diversity and transfer of *Isa* resistance genes in group B *Streptococcus*. *J Antimicrob Chemother*. 2015;dkv277. <https://doi.org/10.1093/jac/dkv277>.
- Rocco JM, Churchward G. The integrase of the conjugative transposon Tn916 directs strand- and sequence-specific cleavage of the origin of conjugal transfer, oriT, by the endonuclease Orf20. *J Bacteriol*. 2006;188:2207–13. <https://doi.org/10.1128/JB.188.6.2207-2213.2006>.
- Wright LD, Grossman AD. Autonomous replication of the conjugative transposon Tn916. Zhulin IB, editor. *J Bacteriol*. 2016;198:3355–66. <https://doi.org/10.1128/JB.00639-16>.
- Lee CA, Grossman AD. Identification of the origin of transfer (*oriT*) and DNA relaxase required for conjugation of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *J Bacteriol*. 2007;189:7254–61. <https://doi.org/10.1128/JB.00932-07>.
- Carr SB, Phillips SEV, Thomas CD. Structures of replication initiation proteins from staphylococcal antibiotic resistance plasmids reveal protein asymmetry and flexibility are necessary for replication. *Nucleic Acids Res*. 2016;44:2417–28. <https://doi.org/10.1093/nar/gkv1539>.
- Nash RP, Habibi S, Cheng Y, Lujan SA, Redinbo MR. The mechanism and control of DNA transfer by the conjugative relaxase of resistance plasmid pCU1. *Nucleic Acids Res*. 2010;38:5929–43. <https://doi.org/10.1093/nar/gkq303>.
- Datta S, Larkin C, Schildbach JF. Structural insights into single-stranded DNA binding and cleavage by F factor Tral. *Structure*. 2003;11:1369–79. <https://doi.org/10.1016/j.str.2003.10.001>.

37. Ilangovan A, Kay CWM, Roier S, El Mkami H, Salvadori E, Zechner EL, et al. Cryo-EM structure of a Relaxase reveals the molecular basis of DNA unwinding during bacterial conjugation. *Cell*. 2017;169:708–721.e12. <https://doi.org/10.1016/j.cell.2017.04.010>.
38. Edwards JS, Betts L, Frazier ML, Pollet RM, Kwong SM, Walton WG, et al. Molecular basis of antibiotic multiresistance transfer in *Staphylococcus aureus*. *Proc Natl Acad Sci*. 2013;110:2804–9. <https://doi.org/10.1073/pnas.1219701110>.
39. Pluta R, Boer DR, Lorenzo-Díaz F, Russi S, Gómez H, Fernández-López C, et al. Structural basis of a histidine-DNA nicking/joining mechanism for gene transfer and promiscuous spread of antibiotic resistance. *Proc Natl Acad Sci*. 2017;114:E6526–35.
40. Chuzeville S, Puymège A, Madec J-Y, Haenni M, Payot S. Characterization of a new CAMP factor carried by an integrative and conjugative element in *Streptococcus agalactiae* and spreading in streptococci. Biswas I, editor. *PLoS One*. 2012;7:e48918. <https://doi.org/10.1371/journal.pone.0048918>.
41. Boyd DA, Mulvey MR. The VanE operon in *Enterococcus faecalis* N00-410 is found on a putative integrative and conjugative element, Tn6202. *J Antimicrob Chemother*. 2013;68:294–9. <https://doi.org/10.1093/jac/dks394>.
42. Han X, Du X-D, Southey L, Bulach DM, Seemann T, Yan X-X, et al. Functional analysis of a bacitracin resistance determinant located on ICE *Cp1*, a novel Tn 916-like element from a conjugative plasmid in *Clostridium perfringens*. *Antimicrob Agents Chemother*. 2015;59:6855–65. <https://doi.org/10.1128/AAC.01643-15>.
43. Machielsens R, Siezen RJ, van Hijum SAFT, van Hylckama Vlieg JET. Molecular description and industrial potential of Tn 6098 conjugative transfer conferring alpha-Galactoside metabolism in *Lactococcus lactis*. *Appl Environ Microbiol*. 2011;77:555–63. <https://doi.org/10.1128/AEM.02283-10>.
44. Rauch PJ, De Vos WM. Characterization of the novel nisin-sucrose conjugative transposon Tn5276 and its insertion in *Lactococcus lactis*. *J Bacteriol*. 1992;174:1280–7. <https://doi.org/10.1128/jb.174.4.1280-1287.1992>.
45. Sansverre EA, Luo X, Park JY, Yoon S, Seo KS, Robinson DA. Transposase-mediated excision, conjugative transfer, and diversity of ICE 6013 elements in *Staphylococcus aureus*. O'Toole G, editor. *J Bacteriol*. 2017;199. <https://doi.org/10.1128/JB.00629-16>.
46. Novick RP. Staphylococcal plasmids and their replication. *Annu Rev Microbiol*. 1989;43:537–65.
47. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–58. <https://doi.org/10.1038/nprot.2015.053>.
48. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010;5:725–38. <https://doi.org/10.1038/nprot.2010.5>.
49. Thomas CD, Balson DF, Shaw WW. *In vitro* studies of the initiation of staphylococcal plasmid replication. *J Biol Chem*. 1990;265:5519–30.
50. Carraro N, Libante V, Morel C, Decaris B, Charron-Bourgoin F, Leblond P, et al. Differential regulation of two closely related integrative and conjugative elements from *Streptococcus thermophilus*. *BMC Microbiol*. 2011;11:238.
51. Bellanger X, Roberts AP, Morel C, Choulet F, Pavlovic G, Mullany P, et al. Conjugative transfer of the integrative conjugative elements ICES<sub>t1</sub> and ICES<sub>t3</sub> from *Streptococcus thermophilus*. *J Bacteriol*. 2009;191:2764–75. <https://doi.org/10.1128/JB.01412-08>.
52. Puymège A, Bertin S, Guédon G, Payot S. Analysis of *Streptococcus agalactiae* pan-genome for prevalence, diversity and functionality of integrative and conjugative or mobilizable elements integrated in the tRNA<sub>Lys</sub> CTT gene. *Mol Gen Genomics*. 2015;290:1727–40. <https://doi.org/10.1007/s00438-015-1031-9>.
53. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. 2007;1:2876–90. <https://doi.org/10.1038/nprot.2006.202>.
54. Lorenzo-Díaz F, Dostal L, Coll M, Schildbach JF, Menendez M, Espinosa M. The MobM relaxase domain of plasmid pMV158: thermal stability and activity upon Mn<sup>2+</sup> and specific DNA binding. *Nucleic Acids Res*. 2011;39:4315–29. <https://doi.org/10.1093/nar/gkr049>.
55. Kopec J, Bergmann A, Fritz G, Grohmann E, Keller W. TraA and its N-terminal relaxase domain of the gram-positive plasmid pIP501 show specific *oriT* binding and behave as dimers in solution. *Biochem J*. 2005;387:401–9.
56. Guasch A, Lucas M, Moncalian G, Cabezas M, Pérez-Luque R, Gomis-Rüth FX, et al. Recognition and processing of the origin of transfer DNA by relaxase TrwC. *Nat Struct Biol*. 2003;10:1002–10.
57. Monzingo AF, Ozburn A, Xia S, Meyer RJ, Robertus JD. The structure of the minimal Relaxase domain of MobA at 2.1 Å resolution. *J Mol Biol*. 2007;366:165–78. <https://doi.org/10.1016/j.jmb.2006.11.031>.
58. Koepsel RR, Murray RW, Rosenblum WD, Khan SA. The replication initiator protein of plasmid pT181 has sequence-specific endonuclease and topoisomerase-like activities. *Proc Natl Acad Sci*. 1985;82:6845–9.
59. Lee CA, Babic A, Grossman AD. Autonomous plasmid-like replication of a conjugative transposon. *Mol Microbiol*. 2010;75:268–79. <https://doi.org/10.1111/j.1365-2958.2009.06985.x>.
60. Guzmán LM, Espinosa M. The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid *oriT*. *J Mol Biol*. 1997;266:688–702.
61. Llosa M, Grandoso G, Hernando MA, de la Cruz F. Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function. *J Mol Biol*. 1996;264:56–67.
62. Lanka E, Wilkins BM. DNA processing reactions in bacterial conjugation. *Annu Rev Biochem*. 1995;64:141–69.
63. Xia S, Robertus JD. Effect of divalent ions on the minimal relaxase domain of MobA. *Arch Biochem Biophys*. 2009;488:42–7. <https://doi.org/10.1016/j.abb.2009.06.004>.
64. Larkin C, Haft RJF, Harley MJ, Traxler B, Schildbach JF. Roles of active site residues and the HUH motif of the F plasmid *Tral* relaxase. *J Biol Chem*. 2007;282:33707–13. <https://doi.org/10.1074/jbc.M703210200>.
65. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*. 1988;16:10881–90.
66. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21. <https://doi.org/10.1093/sysbio/syq010>.
67. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. 1992;8:275–82.
68. Geer LY. CDART: protein homology by domain architecture. *Genome Res*. 2002;12:1619–23. <https://doi.org/10.1101/gr.278202>.
69. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26:680–2. <https://doi.org/10.1093/bioinformatics/btq003>.
70. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf*. 2004;5:113.
71. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673–80.
72. Crooks GE. WebLogo: A Sequence Logo Generator. *Genome Res*. 2004;14:1188–90. <https://doi.org/10.1101/gr.849004>.
73. Maguin E, Duwat P, Hege T, Ehrlich D, Gruss A. New thermosensitive plasmid for gram-positive bacteria. *J Bacteriol*. 1992;174:5633–8.
74. Biswas I, Gruss A, Ehrlich SD, Maguin E. High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J Bacteriol*. 1993;175:3628–35.
75. Que Y-A, Haefliger J-A, Francioli P, Moreillon P. Expression of *Staphylococcus aureus* clumping factor a in *Lactococcus lactis* subsp. *cremoris* using a new shuttle vector. *Infect Immun*. 2000;68:3516–22.
76. Dahmane N, Libante V, Charron-Bourgoin F, Guédon E, Guédon G, Leblond-Bourget N, et al. Diversity of integrative and conjugative elements of *Streptococcus salivarius* and their intra- and interspecies transfer. *Appl Environ Microbiol*. 2017;83:e00337–17.
77. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000;16:404–5.
78. Adamczak R, Porollo A, Meller J. Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins*. 2005;59:467–75. <https://doi.org/10.1002/prot.20441>.
79. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*. 2015;43:W389–94. <https://doi.org/10.1093/nar/gkv332>.
80. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res*. 2014;42:W337–43. <https://doi.org/10.1093/nar/gku366>.
81. Raghava GPS. APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. *CASP5*. 2002;A-132.

82. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem.* 2000;287:252–60. <https://doi.org/10.1006/abio.2000.4880>.
83. Whitmore L, Wallace BA. DICROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 2004;32:W668–73. <https://doi.org/10.1093/nar/gkh371>.
84. Böhm G, Muhr R, Jaenicke R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng Des Sel.* 1992;5:191–5.
85. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER: Comparative Protein Structure Modeling Using Modeller. In: Bateman A, Pearson WR, Stein LD, Stormo GD, Yates JR, editors. *Current Protocols in Bioinformatics*. Hoboken: Wiley; 2016. p. 5.6.1–5.6.37. <https://doi.org/10.1002/cpbi.3>.
86. Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 2006;15:2507–24. <https://doi.org/10.1110/ps.062416606>.
87. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005;26:1781–802. <https://doi.org/10.1002/jcc.20289>.
88. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem.* 2013;34: 2135–45. <https://doi.org/10.1002/jcc.23354>.
89. Pavlovic G, Burrus V, Gintz B, Decaris B, Guédon G. Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICES $t1$ -related elements from *Streptococcus thermophilus*. *Microbiology.* 2004;150:759–74. <https://doi.org/10.1099/mic.0.26883-0>.
90. Nakayama N, Narumi I, Nakamoto S, Kihara H. Complete nucleotide sequence of pSTK1, a cryptic plasmid from *Bacillus stearothermophilus* TK015. *Biotechnol Lett.* 1993;15:1013–6.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

